

EXHIBIT DD

JOSEPH SAVERI
LAW FIRM

601 CALIFORNIA STREET
SUITE 1505
SAN FRANCISCO CA 94108

TEL 415.500.6800
FAX 415.395.9940

August 22, 2024

Via Electronic Mail

Judd Lauter
Cooley LLP
3 Embarcadero Center, 20th Floor
San Francisco, CA 94111
Email: jlauter@cooley.com

Re: *Kadrey et al, v. Meta Platforms, Inc*, Case No. 3:23-cv-03417-VC
(N.D. Cal): Deficiencies in Meta's Production and Source Code
Inspection

Judd:

We write on behalf of Plaintiffs regarding Meta Platforms Inc.'s ("Meta") productions in response to Plaintiffs' Requests for Productions ("RFPs"). Specifically, we write regarding the following deficiencies and other issues including Meta's: (I) deficient production to date regarding the specific RFPs listed below; (II) production of documents that are cropped and therefore only partially readable; (III) redactions on communications not involving attorneys; (IV) production of documents containing hyperlinks to other internal documents that Plaintiffs do not currently have access to; (V) production of documents containing text in a foreign language; and (VI) issues pertaining to additional miscellaneous items.

We also write regarding deficiencies in the source code inspection that occurred at your law firm's Palo Alto offices during the week of August 12. *See* section VII below.

By addressing these deficiencies, Plaintiffs do not waive their right to address additional deficiencies in Meta's productions as we continue to review the productions.

I. Deficient Production of Documents Related to Specific RFPs

Training data. As you know, Plaintiffs' RFPs request the Training Data for Llama 1, Llama 2, and Llama 3. *See* RFPs 1, 2, and 3. It appears Meta

August 22, 2024

Page 2

has produced only the training data for Llama 1. Please produce the training data for Llama 2 and Llama 3.

Communications with Shadow Libraries. We served RFPs for Meta's communications with Library Genesis, Z-Library, Sci-Hub, Bibliotik, The Eye, and Anna's Archive regarding training data. *See* RFPs 7-12. Meta's production does not contain any such documents. Is it Meta's position that no such documents exist?

Communications with Certain Business Partners. We served RFPs for Meta's communications with Microsoft Corporation, OpenAI, Anthropic, Dell Inc., Cloudflare, Inc. and Qualcomm regarding training data. *See* RFPs 29-34. We did not encounter any such documents in Meta's production. We find it difficult to believe that no such documents would exist, given Meta's partnerships with these organizations. Please produce them.

Additional Documents re Licensing Training Data (RFP 45). Our review of the documents produced indicates that Meta created a process for identifying potential licensors and reaching out to "BD partners [to] ask them if they have any leads on entities who may have this kind of data." We have not seen any documents or communications between Meta researchers and BD partners regarding potential licensing. Please produce those documents. Similarly, Meta produced a document entitled "Generative AI BD Pipeline – 2023,"¹ which would appear to assign BD partners to the task of reaching out to specific content companies; however, Meta has not produced any information regarding those reachouts. Such communications and any internal notes or memoranda regarding same would be responsive to RFP 45 and they are highly relevant to this case. Please produce them. Plaintiffs also believe the BD partners listed in the aforementioned document should be designated as additional document custodians.

Actual or Projected Income (RFP 46) and Financials (RFP 53). Meta's production is also lacking in that there do not appear to be any documents regarding Meta's actual or projected income from the sale or licensing of the Meta Language Models (RFP 46), and Meta did not produce the financial documents requested in RFP No. 53. Is it Meta's position that such documents do not exist? This would be surprising.

¹ *See* Meta_Kadrey_00045315.

August 22, 2024

Page 3

Documents and Communications Concerning individuals or entities that have been given access to, or denied or limited access from Llama 1 or Llama 2 (RFP 50). Meta has not produced any such documents.

II. Production of Documents that are Cropped and Therefore Only Partially Readable

Many of the documents in Meta's production are only partially readable because portions of those documents are partially cropped out, perhaps due to a technical issue in the production.

Examples of these documents include, without limitation, documents with the following beginning bates labels:

- Meta_Kadrey_00032830
- Meta_Kadrey_00033593
- Meta_Kadrey_00034091
- Meta_Kadrey_00035495
- Meta_Kadrey_00048940
- Meta_Kadrey_00052504
- Meta_Kadrey_00056518
- Meta_Kadrey_00057081
- Meta_Kadrey_00065244
- Meta_Kadrey_00065631
- Meta_Kadrey_00089198

The issue described above is pervasive throughout Meta's production; the above-listed documents are but a few relevant examples. Please reproduce all documents in Meta's production that have this issue.

III. Meta's Redactions

Meta has produced hundreds of documents with redactions, and we look forward to receiving a privilege log with entries for each redaction in Meta's production. In an email from Ms. Stameshkin dated August 9, Meta indicated it would produce a privilege log by August 30.

It would appear that Meta has redacted documents on the basis of relevance or confidentiality. In this regard, Plaintiffs note that Meta produced several communications containing redactions even though there appear to be no attorneys on the document. (See, e.g., Meta_Kadrey_00074729, Meta_Kadrey_00088838, Meta_Kadrey_00088842.) To the extent that Meta has redacted on grounds other than privilege, such redactions would be improper. In fact, Meta took the position that documents should not be redacted on

August 22, 2024

Page 4

relevancy grounds, and Plaintiffs promptly reproduced the at-issue documents without redactions. We'd request the same from Meta.

IV. Documents Containing Hyperlinks

Several of the Documents produced by Meta contain hyperlinks to documents that apparently exist on the cloud, but are currently not accessible to Plaintiffs. Although our review is still underway, we have identified the following such documents:

- Meta_Kadrey_00065246 (beginning bates: Meta_Kadrey_00065244)
- Meta_Kadrey_0006334 (beginning bates: Meta_Kadrey_00006135)
- Meta_Kadrey_00074357
- Meta_Kadrey_00074359
- Meta_Kadrey_00079256
- Meta_Kadrey_00090897

Please produce the documents associated with hyperlinks contained in the above-identified documents.

V. Foreign Language Documents

Upon reviewing Meta's production, Plaintiffs discovered that a significant quantity of documents (at least 1,463 documents) contain writings in French. Plaintiffs are proceeding with the retention of a certified translator, although some advance notice by Defendants would have been useful, especially in light of the looming fact discovery cut-off and the fact that Plaintiffs granted Meta's request for an extension on the substantial completion deadline. Meta's failure to do so has caused unnecessary delay.

VI. Miscellaneous Issues with Meta's Document Production

On August 9, we sent you an email requesting the surnames of the individuals appearing on the charts in Meta_Kadrey_00045315-337. We sent a follow up email on August 16. We have not heard from you. Please provide those as soon as possible. As noted above, we believe these individuals should be designated as additional custodians.

Additionally, the document Bates labeled Meta_Kadrey_00078992 states: "In general, the recommendation aligned with deleting some of the problematic data." Was anything destroyed pursuant to this directive? Please provide us with an answer by August 27.

August 22, 2024

Page 5

VII. Issues with Source Code Inspection

Absence of repository metadata. According to the source code on the inspection machine, the canonical version of the code is apparently hosted at <https://github.com/fairinternal> (and maybe other locations on GitHub not apparent from the inspection). A repository hosted on GitHub (or Gitlab, or other web-based source-hosting service) includes a lot of important metadata, including commit history (i.e., a time series of changes made to the code, often with descriptive comments), issues (i.e., discussion threads where software contributors discuss code-related topics), and pull requests (i.e., proposed changes to the code, which may also have discussions and code reviews associated with them). All this metadata is responsive to Plaintiffs' discovery requests, but none of it appears on the inspection machine. The repository metadata for each applicable repository needs to be either produced or otherwise made inspectable. Until then, Meta's production is insufficient.

Inability to retrieve or print digital notes. The protective order allows Plaintiffs to ask for files on the inspection machine to be produced. The protective order also forbids Plaintiffs from using digital devices in the inspection room (other than the inspection computer). Taken together, this means that Plaintiffs must manually write down the path of every file sought. Consistent with the protective order, Plaintiffs are using computational tools on the inspection machine to identify files of interest and create reports of files sought. It is unreasonable to insist that Plaintiffs manually copy scores of file paths by hand. This is time-consuming, error-prone, and also makes more work for Meta (who must then translate the handwritten files back to digital).

Plaintiffs propose the following amendment to the protective order to facilitate digital notes:

DIGITAL NOTES. The Inspecting Party's counsel and/or experts may create and store electronic notes on the secured computers in scratch files. The Inspecting Party will not waive any applicable work-product protection over their electronic notes by saving them to the secured computers temporarily. Any notes so stored will be treated as "HIGHLY CONFIDENTIAL – ATTORNEYS' EYES ONLY." The Inspecting Party may obtain productions of electronic notes taken on the secure computer. To make a request, the Inspecting Party shall create a directory entitled "Production Request" and save the notes in that directory. The Inspecting Party shall alert Meta when it has saved the notes in the "Production Request" directory

August 22, 2024

Page 6

that it requests to be produced. Upon receiving a request, Meta shall Bates number, and label ‘HIGHLY CONFIDENTIAL – ATTORNEYS’ EYES ONLY’” all requested pages, and produce them to the Inspecting Party.

Dataset Explorer. Plaintiffs have identified through Meta’s production that there exists a tool that allows filtering of training data (which Meta previously told us does not exist). *See* Meta_Kadrey_00056492-500. Plaintiffs require direct access to the dataset explorer from the inspection for our next inspection. Plaintiffs believe this the most efficient path forward.

Please let us know when you are available for a meet and confer on August 26 or August 27 to discuss the foregoing issues.

Sincerely,

/s/ Holden Benon

Holden Benon

cc: All counsel of record (via email)